Tema 2. Estimació puntual i per interval

2.1. Objecte de l'estimació estadística

Per què estimem? Com estimem?

• A EEiE_I podíem fer el càlcul de probabilitats atès que es coneixien els paràmetres poblacionals que acabaven de definir/concretar una distribució poblacional, però:

- Com sabem si, per exemple, una determinada població segueix una N(2,3)?
- Qui i com ha dit que $\mu=2$ o que $\sigma^2=9$?
- Per què sabem que és una normal? I si és una Poisson? Quin seria el paràmetre λ?

• Aquests valors són desconeguts i per tant serà necessari ESTIMAR AQUESTS PARÀMETRES. Disposar d'estimacions correctes per a prendre decisions i això es farà fent ús d'ESTIMADORS el més adients possibles

• Això vol que hi ha més d'un estimador possible? Com escollir entre vàries propostes d'estimadors?

Propietats dels estimadors
- • És un estimador NO ESBIAIXAT?
- • És un estimador EFICIENT?
- • És un estimador CONSISTENT?

La resposta a aquestes preguntes es permetran escollir entres diverses propostes d'estimadors, aquell que s'adapti millor a les nostres necessitats

Tema 2. Estimació puntual i per interval

• En determinants casos serà MOLT EVIDENT quin és el millor ESTIMADOR per a obtenir una ESTIMACIÓ D'UN DETERMINAT PARÀMETRE POBLACIONAL, però en altres casos NO.
• Quan no està clar quin estimador proposar, tenim instruments per a poder GENERAR estimadors en sintonia amb els paràmetres poblacionals que volem estimar?

Mètodes d'estimació
- Pel mètode dels moments
- Per màxima versemblança

• Un cop hem seleccionat entre els possibles estimadors aquell que considerem més correcte

• És exactament igual al paràmetre poblacional???
• Com podem saber si és igual a quelcom que desconeixem????
• Com podem valorar si hem comés un error si aquest error és la diferència entre la nostra estimació i quelcom desconegut i, per tant, es un error desconegut?

Dues possibilitats:
- • Fixar un error màxim a cometre i que considerem adient assumir (Últim punt Tema 2)
- • Assignar una probabilitat a l'error que podem cometre (Resta Tema 2)

Tema 2. Estimació puntual i per interval

• Aquesta assignació de probabilitat ens permetrà fer UNA ESTIMACIÓ PER INTERVAL. És a dir, assignar una probabilitat al fet que el PARÀMETRE POBLACIONAL es trobi dins determinats valors com, per exemple:

P(Límit Inferior – PARÀMETRE – Límit Superior)= 0.95

2.2. Definició i característiques d'un estimador

DEFINICIÓ: Un estimador d'un paràmetre poblacional θ, que simbolitzarem com $\hat{\theta}$ és una funció de les observacions mostrals que es fan servir per a generar estimacions dels valor del paràmetre poblacional $\hat{\theta} = g(X_1, X_2, \cdots, X_n)$

• Al ser una funció de variables aleatòries ÉS UNA VARIABLE ALEATÒRIA.

• Quin és el principal problema de l'estimació? Que combinacions dels elements d'una mostra hi ha moltes. Exemple:

Sigui X la durada (en hores) de la bateria d'un mòbil en situació d'espera

X_1=130 X_2=142 X_3=145 X_4=153 X_5=158 X_6=159 X_7= 160 X_8=162 X_9=165 X_{10}=171

Es pretén estimar μ mitjançant els següents estimadors:

$\hat{\mu}_1 = \overline{X}$ $\quad\hat{\mu}_2 =$ Mediana $\quad\hat{\mu}_3 = \overline{X}_6$ (Mitjana dels 6 valors centrals)

Obtenint les estimacions $\hat{\mu}_1 = 154{,}5$ $\quad\hat{\mu}_2 = 158{,}2$ $\quad\hat{\mu}_3 = 156{,}2$ QUINA ÉS LA MILLOR?

Tema 2. Estimació puntual i per interval

• IDEA: analitzar les propietats dels estimadors per a respondre a la següent pregunta.

Si traguéssim mostres successives, quina d'elles generaria ESTIMACIONS MÉS PROPERES AL VERITABLE VALOR DEL PARÀMETRE POBLACIONAL???

2.3. Error Quadràtic Mitjà: Característiques dels estimadors

Quines són les característiques que seria desitjable que un BON ESTIMADOR PRESENTÉS?

1.- NO ESBIAIXAMENT

BIAIX: És la diferència entre el valor esperat i el veritable valor poblacional del paràmetre (DESCONEGUT)

$$BIAIX(\theta) = E(\hat{\theta}) - \theta$$

Si $BIAIX(\theta) > 0 \quad (< 0)$, l'estimador tendeix sistemàticament a generar estimacions per sobre (sota) del valor poblacional del paràmetre. És a dir, sobreestima (subestima).

ESTIMADOR NO ESBIAIXAT (CENTRAT). És aquell estimador amb BIAIX NUL. És a dir, la seva esperança matemàtica coincideix amb el valor poblacional del paràmetre:

$$E(\hat{\theta}) = \theta \quad \Rightarrow \quad BIAIX(\theta) = 0$$

Tema 2. Estimació puntual i per interval

• IMPORTANT: Un estimador NO ESBIAIXAT ens està informant què amb determinades mostres generarà valors per sobre del paràmetre poblacional, en canvi, amb d'altres generarà valors per sota. És a dir, NO TÉ UN COMPORTAMENT SISTEMÀTIC, ni en un sentit ni en l'altre. EXEMPLE:

Es proposa $\overline{X}$ com estimador de μ. Sabem que $E(\overline{X}) = \mu$, per tant, $\overline{X}$ és un estimador NO ESBIAIXAT de μ

Si es proposa $S_*^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ com estimador de σ^2 sabem que en aquest cas també és un estimador NO ESBIAIXAT de σ^2, atès que $E(S_*^2) = \sigma^2$

En canvi, si es proposa com estimador $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$, es pot demostrar que $E(\hat{\sigma}^2) \neq \sigma^2$, de fet, és: $E(\hat{\sigma}^2) = \dfrac{n-1}{n}\sigma^2$, i per tant, el

$$BIAIX(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-1}{n}\sigma^2$$

• El NO ESBIAIXAMENT és una condició NECESSÀRIA, però NO SUFICIENT.

ECET 2.1

Tema 2. Estimació puntual i per interval

2.- <u>EFICIÈNCIA:</u> Sembla raonable què, entre tots els estimadors NO ESBIAIXATS, escollim aquell que presenti una menor variabilitat.

$\hat{\theta}_i$ és un estimador més eficient que $\hat{\theta}_j$ si:

1er. És NO ESBIAIXAT

2on. $Var(\hat{\theta}_i) \leq Var(\hat{\theta}_j) \quad \forall i \neq j$

Per tant, l'ESTIMADOR MÉS EFICIENT, serà aquell que presenti una VARIÀNCIA MÍNIMA. Però podem saber de tots els estimadors possibles quin és el que presenta la variància mínima?.

EFICIÈNCIA ABSOLUTA Vs. EFICIÈNCIA RELATIVA

Existeix un mecanisme que permet trobat quina és la variància mínima. Un estimador és EFICIENT EN TERMES ABSOLUTS si assoleix la COTA DE CRAMER-RAO. Així, si la Variància d'un estimador assoleix aquesta COTA, sabrem que és l'ESTIMADOR DE VARIÀNCIA MÍNIMA i, per tant, EL MÉS EFICIENT (EFICIÈNCIA ABSOLUTA). És a dir:

$$Var(\hat{\theta}) \geq \frac{\left[1 - BIAIX'(\hat{\theta})\right]^2}{nE\left(\dfrac{\delta \ln f(X;\theta)}{\delta\theta}\right)^2}$$

Ara bé, podria donar-se el cas que cap dels estimadors seleccionats arribés a assolir la COTA de CRAMER-RAO

EFICIÈNCIA RELATIVA

Tema 2. Estimació puntual i per interval

3.- <u>EFICIÈNCIA RELATIVA:</u>

Siguin $\hat{\theta}_1$ i $\hat{\theta}_2$ dos estimadors NO ESBIAIXATS de θ, diem que $\hat{\theta}_1$ és més eficient que $\hat{\theta}_2$ si $V(\hat{\theta}_1) < V(\hat{\theta}_2)$. Per tant, l'EFICIÈNCIA RELATIVA, es determina segons:

$$\frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} \begin{cases} >1 & \hat{\theta}_1 \text{ MÉS EFICIENT QUE } \hat{\theta}_2 \\ <1 & \hat{\theta}_1 \text{ MENYS EFICIENT QUE } \hat{\theta}_2 \\ =1 & \hat{\theta}_1 \text{ IGUAL EFICIENT QUE } \hat{\theta}_2 \end{cases}$$

4.- <u>FORMA FUNCIONAL:</u> Aquesta característica fa referència a la forma funcional en que combinem els elements de la mostra. Així, sembla preferible que aquesta forma funcional sigui el més senzilla possible. Sigui $\hat{\theta}_1 = g(X_1, X_2, \cdots, X_n)$. f(·) és una funció lineal, $\hat{\theta}_1$ és ESTIMADOR LINIAL de θ

ESTIMADOR LINEAL NO ESBIAIXAT ÒPTIM (BLUE: Best Lineal Unbiased Estimattor) (ELIO: Estimador Linial Insesgado Óptimo):

1er. $\hat{\theta}_1 = f(X_1, X_2, \cdots X_n)$ on f(·) és una funció lineal

2on. $E(\hat{\theta}_1) = \theta$

3er. $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_i)$ on $\hat{\theta}_i$ és qualsevol ESTIMADOR LINIEAL NO ESBIAIXAT de θ

Exemple DIANA. Existeix alguna forma de tenir present les dues característiques. NO BIAIX i EFICIÈNCIA??

Tema 2. Estimació puntual i per interval

5.- <u>ERROR QUADRÀTIC MITJÀ:</u> L'EQM mesura la dispersió de l'ESTIMADOR respecte el veritable valor de la PARÀMETRE poblacional.

$$EQM(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

Valors baixos → Dispersió petita
Valors alts → Dispersió alta

S'ha de diferenciar respecte a la VARIÀNCIA (Dispersió respecte el valor esperat).

$$Var(\hat{\theta}) = E(\hat{\theta} - E(\theta))^2$$

$$EQM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = E(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2 =$$

$$= E\left((\hat{\theta} - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \theta)^2 + 2(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) \right) =$$

$$= E\left[(\hat{\theta} - E(\hat{\theta}))^2 \right] + E\left[(E(\hat{\theta}) - \theta)^2 \right] + 0 =$$

$$= Var(\hat{\theta}) + BIAIX^2(\hat{\theta})$$

1.- L'EQM depèn de la variància i del BIAIX
2º.- Si l'estimador és NO ESBIAIXAT l'EQM és igual a la Variància de l'estimador.

6.- <u>CARACTERÍSTIQUES ASIMPTÒTIQUES:</u> Fan referència a com es comporta la DISTRIBUCIÓ DE PROBABILITAT DE L'ESTIMADOR quan la mida de la mostra tendeix a INFINIT.

A.) NO ESBIAIXAMENT ASIMPTÒTIC. Un ESTIMADOR és NO ESBIAIXAT ASIMPTÒTICAMENT si quan n → ∞ el seu BIAIX tendeix a 0.

$$\lim_{n\to\infty}\left[BIAIX\left(\hat{\theta}\right)\right]=0 \quad \Rightarrow \quad \lim_{n\to\infty}\left[E\left(\hat{\theta}\right)\right]=\theta$$

B.) CONSISTÈNCIA: Valora la forma en que la dispersió de l'estimador respecte al veritable valor del paràmetre poblacional es comporta quan n →∞. Existeixen 3 tipus de consistència:

Dèbil → b.1.- Consistència en Probabilitat → $\lim_{n\to\infty} P\left[\left|\hat{\theta}-\theta\right|\geq\varepsilon\right]=0$

Forta:
- b.2.- Consistència Quasi Segura. → $P\left(\lim_{n\to\infty}\hat{\theta}=\theta\right)=1$
- b.3.- Consistència en EQM → $\lim_{n\to\infty}\left[EQM\left(\hat{\theta}\right)\right]=\lim_{n\to\infty}\left[Var\left(\hat{\theta}\right)\right]+\lim_{n\to\infty}\left[BIAIX^2\left(\hat{\theta}\right)\right]=0$

1.- Per a que un ESTIMADOR NO ESBIAIXAT sigui CONSISTENT:

$$\lim_{n\to\infty}\left[EQM\left(\hat{\theta}\right)\right]=\lim_{n\to\infty}\left[Var\left(\hat{\theta}\right)\right]=0$$

2º.- Per a que un ESTIMADOR ESBIAIXAT sigui consisten:

$$\lim_{n\to\infty}\left[Var\left(\hat{\theta}\right)\right]=0 \quad \text{i} \quad \lim_{n\to\infty}\left[BIAIX^2\left(\hat{\theta}\right)\right]=0$$

ECET 2.2 01_T2_Propietats_Estimadors.R

<u>2.4.- Mètodes d'estimació: Mètode dels Moments i Mètode de Màxima Versemblança.</u>

A.) MÈTODE DELS MOMENTS (Pearson 1856-1936). És el mètode d'estimació més senzill i es basa en la idea de similitud entre els moments potencials de dos variables aleatòries. És a dir els PARÀMETRES POBLACIONALS es poden estimar a partir dels Moments potencials de la mostra.

A.1) La distribució de la mostra aleatòria depèn d'un ÚNIC PARÀMETRE θ

1er.- Especifiquem la relació que Existeix entre el Primer moment poblacional i el paràmetre θ. → $\mu = g(\theta)$

2on.- Substituïm el moment poblacional pel mostral, → $\overline{X} = g\left(\hat{\theta}\right)$

3er.- Aïllem l'ESTIMADOR en funció del primer moment mostral. → $\hat{\theta} = f\left(\overline{X}\right)$

A.2) Si la distribució depèn de més paràmetres $(\theta_1,\theta_2,\ldots,\theta_n)$ s'hauran d'igualar tants moments com paràmetres i resoldre i SISTEMA D'EQUACIONS.

NOTA: S'ha de tenir present que el mètode no te en compte la distribució de probabilitat mostral i, per tant, no es garanteixen moltes propietats.

PROPIETATS: 1ª.- Si s'obtenen a partir del PRIMER MOMENT ORDINARI els estimadors obtinguts són NO ESBIAIXATS. 2ª.- Sota condicions molt generals són CONSISTENTS en EQM. 3ª.- Presenten NORMALITAT ASIMPTÒTICA, és a dir, quan n → ∞ la distribució de l'estimador tendeix a una normal.

ECET 2.3

Tema 2. Estimació puntual i per interval

B.) MÈTODE DE MÀXIMA VERSEMBLANÇA.

• És una dels mètodes més utilitzats per a ESTIMACIÓ PUNTUAL.

• Es tracta d'escollir aquell valor del paràmetre que maximitzi la versemblança que la població hagi generat la nostra mostra.

• Sabem que una població pot generar diferents mostres, però és que a més, una mostra pot haver estat generada per diferents poblacions.

Suposem que X segueix un N(μ,2) on μ és desconeguda i disposem d'una mostra de mida n (X_1, X_2,…,X_n). Quina població ha generat aquesta mostra?

Una N(μ_1,2)?, una N(μ_2,2)? o una N(μ_k,2)? → INFINITES POSSIBILITATS.

• Per a obtenir l'ESTIMADOR MAXIMVERSEMBLANT s'ha de:

1er.- Especificar la funció de Versemblança L(X,θ)

2on.- Determinar el màxim igualant la primera derivada respecte el paràmetre a cero.

3er.- Comprovar que la segona derivada és negativa.

• Si només existeix una paràmetre a estimar NOMÉS existirà una primera derivada.

• Si la funció depèn de k paràmetres existiran tantes derivades parcials com paràmetres i, en conseqüència s'hauran de resoldre el conseqüent SISTEMA D'EQUACIONS.

Tema 2. Estimació puntual i per interval

LLEIS:

• LLEI DE VERSEMBLANÇA: Una mostra informa millor sobre un PARÀMETRE θ_1 que sobre un altre θ_2 si la VERSEMBLANÇA del primer és més gran que la del segon.

•RAÓ DE VERSEMBLANÇA: La informació continguda a la mostra permetrà escollir entre dos paràmetres mitjançant el quocient de versemblança

$$\lambda(X) = \frac{L(X,\theta_1)}{L(X,\theta_2)}$$

PROPIETATS:

1ª.- En general, són estimadors NO ESBIAIXATS i si no ho són, són ASIMPTÒTICAMENT NO ESBIAIXATS.

2ª.- Sota certes condicions són CONSISTENTS.

3ª.- Si existeix un estimador amb variància igual a la COTA DE CRAMER-RAO, aquest és l'estimador màximversemblant. No tot estimador màximversemblant és eficient, malgrat, però, si existeix aquest estimador estimador eficient, aquest és màximversemblant.

4ª.- Presenten normalitat i eficiència asimptòtiques.

5ª.- Son invariants. És a dir si es fan transformacions del paràmetre poblacional la mateixa transformació de l'estimador és un estimador maximversemblant.

$$\begin{array}{ccc} \hat{\theta} & \rightarrow MV \rightarrow & \theta \\ g(\hat{\theta}) & \rightarrow MV \rightarrow & g(\theta) \end{array}$$

ECET 2.4

Tema 2. Estimació puntual i per interval

2.5.- Estimació per interval. Interval per a la mitjana, la variància i la proporció poblacional. Intervals asimptòtics.

Fins ara hem pogut estimar θ de forma que es verifiquin totes les propietats que serien desitjables en un estimador, però quan particularitzem aquesta estimació en el valor numèric concret (ESTIMACIÓ PUNTUAL) per a una mostra determinada. Desconeixem si aquest valor està pròxim a θ. Degut, precisament, a l'aleatorietat de la mostra. Per això, s'han d'instrumentalitzar mesures que permetin calibrar el possible error d'una estimació puntual.

Així, s'acostuma a acompanyar una ESTIMACIÓ PUNTUAL amb un INTERVAL, on es CONFIA que es trobi el VERITABLE PARÀMETRE POBLACIONAL amb un determinat MARGE D'ERROR.

INTERVAL DE CONFIANÇA: Conjunt de valors que amb una DETERMINADA PROBABILITAT (GRAU DE CONFIANÇA) conté el veritable valor del paràmetre que es desitja estimar.

El primer que s'ha de fer es decidir el NIVELL DE CONFIANÇA (1-α), que no és més que la probabilitat que l'interval contingui el paràmetre poblacional:

1er.- El nivell de confiança ha de ser elevat, PERÒ

2on.- Quan més elevat sigui, menys precís serà. $\Rightarrow$ Què fem??

Tema 2. Estimació puntual i per interval

Generalment es treballa amb el 90%, 95% i 99% de NIVELL DE CONFIANÇA (1-α). Amb aquest nivell de confiança trobem els límits inferior ("a") i superior de l'interval ("b") de forma que puguem fer la següent afirmació:

Amb una mostra concreta NO PODEM garantir amb (1-α)*100% de probabilitat que el paràmetre es trobi entre "a" i "b", però si obtinguéssim infinites mostres, esperem que el (1-α)*100% dels intervals creats continguin el veritable valor poblacional del paràmetre. Ara bé, n'hi hauran α*100% d'intervals que no. α rep el nom de

NIVELL DE SIGNIFICACIÓ.

INTERVAL DE CONFIANÇA PER A LA MITJANA POBLACIONAL (UNA MOSTRA)

CAS A VARIANÇA POBLACIONAL CONEGUDA

Sigui una població estadística caracteritzada per $X \approx N(\mu,\sigma)$ on μ es DESCONEGUDA i σ^2 CONEGUDA.

PROCEDIMENT:

1.- Escollim el millor estimador de $\longrightarrow$ $\mu \rightarrow \overline{X}$

2.- Sabem, pel Tema 1 que: $\longrightarrow$ $X \approx N(\mu,\sigma) \rightarrow \overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$

3.- Fixem el nivell de confiança (1-α) que

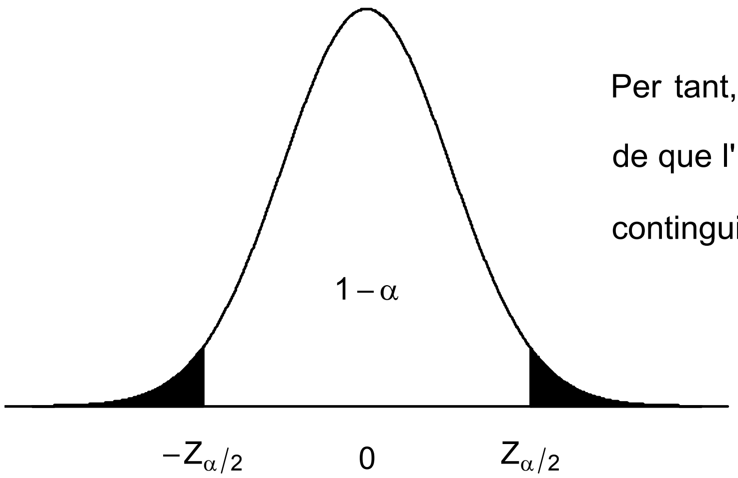
Ens permeti trobar $-Z_{\alpha/2}$ i $Z_{\alpha/2}$ que verifiquin:

$$P\left(-Z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}\right) = 1-\alpha$$

Tema 2. Estimació puntual i per interval

$$P\left(-Z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}\right) = 1-\alpha \quad \Rightarrow \quad P\left(\overline{X} - Z_{\alpha/2}\,\sigma/\sqrt{n} \le \mu \le \overline{X} + Z_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1-\alpha$$

$$P\left(-1.96 \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le 1{,}96\right) = 0.95 \quad \Rightarrow \quad P\left(\overline{X} - 1.96\,\sigma/\sqrt{n} \le \mu \le \overline{X} + 1{,}96\,\sigma/\sqrt{n}\right) = 0.95$$



Per tant, existeix una probabilit at d'1- α de que l'interval aleatòri $\overline{X} \pm Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ contingui el veritable valor de μ

Tema 2. Estimació puntual i per interval

NOTES: $$P\left(\overline{X} - Z_{\alpha/2}\,\sigma/\sqrt{n} \le \mu \le \overline{X} + Z_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1-\alpha$$

1.- L'interval està centrat en la mitjana mostral:

2.- L'amplitud de l'interval es $2 * Z_{\alpha/2}\,\sigma/\sqrt{n}$

I depèn de:

2.1 Quan més gran sigui la variància poblacional major serà l'amplitud (menys precisió)

2.2. És invers a la mida mostral. Quan més gran sigui aquesta menor serà l'amplitud. (major precisió)

2.3. Quan més gran sigui el nivell de confiança major serà l'amplitud (menor precisió).

3.- Es denomina error màxim d'estimació a : $d = Z_{\alpha/2}\,\sigma/\sqrt{n}$

4.- L'interval és aplicable quan la variable no es distribueix segons una normal per n>30 (TCL)

06_T2_Interval_MitjanaPoblacional_cadamostra_CON.R

06_T2_Interval_MitjanaPoblacional_valoresperat_CON.R

ECET 2.5

Tema 2. Estimació puntual i per interval

INTERVAL DE CONFIANÇA PER A LA MITJANA POBLACIONAL (UNA MOSTRA)

CAS B VARIANÇA POBLACIONAL DESCONEGUDA

Sigui una població estadística caracteritzada per $X \approx N(\mu,\sigma)$ on μ i σ^2 són DESCONEGUDES. A la pràctica, si es desconex μ és molt probable que σ també es desconegui

PRODEDIMENT

1.- Escollim el millor estimador de $\longrightarrow$ $\mu \to \overline{X}$

2.- Sabem, pel Tema 1 que: $\longrightarrow$ $X \approx N(\mu,\sigma) \to \overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Ara bé, com que desconeixem σ^2 hem de fer un estimador de d'aquesta. Aleshores la tipificiació de la variable ja no segueix una distribució N(0,1) Sino una T-STUDENT.

3.- Així, hem de construir un interval que ens permeti trobar $-t_{\alpha/2}$ i $t_{\alpha/2}$ que verifiquin:

$$\begin{cases} Z = \dfrac{\overline{X}-\mu}{S_*/\sqrt{n}} \not\approx N(0,1) \to t_{n-1} \\ P\left(-t_{\alpha/2} \le \dfrac{\overline{X}-\mu}{S_*/\sqrt{n}} \le t_{\alpha/2}\right) = 1-\alpha \end{cases}$$

$$\longrightarrow \quad P\left(\overline{X} - t_{\alpha/2}\, S_*/\sqrt{n} \le \mu \le \overline{X} + t_{\alpha/2}\, S_*/\sqrt{n}\right) = 1-\alpha$$

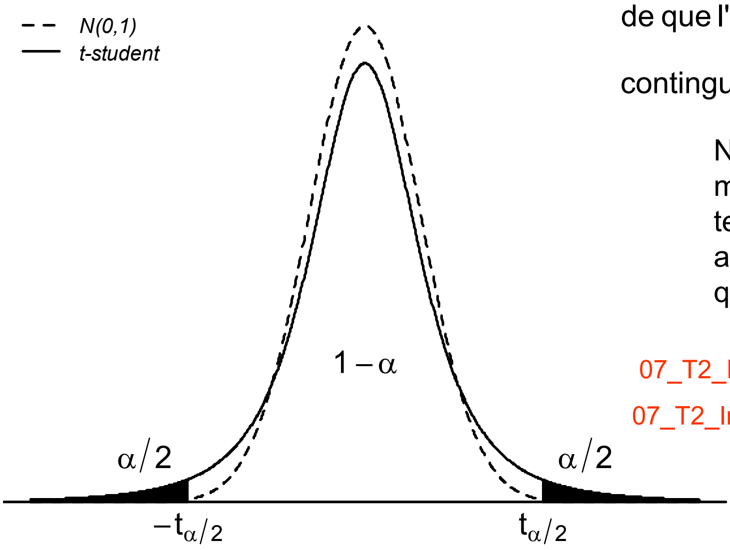
Tema 2. Estimació puntual i per interval

$$P\left(-t_{\alpha/2,n-1} \le \frac{\overline{X}-\mu}{S_*/\sqrt{n}} \le t_{\alpha/2}\right) = 1-\alpha \quad \Rightarrow \quad P\left(\overline{X} - t_{\alpha/2,n-1}\, S_*/\sqrt{n} \le \mu \le \overline{X} + t_{\alpha/2}\, S_*/\sqrt{n}\right) = 1-\alpha$$

Per tant, existeix una probabilitat d'1- α

de que l'interval aleatòri $\overline{X} \pm t_{\alpha/2,n-1} \dfrac{S_*}{\sqrt{n}}$

contingui el veritable valor de μ



NOTA: A mesura que incrementa la mida mostral els valors crítics de la t-Student tendeixen als de una N(0,1). I per això, amb n>30, ja considerem pràcticament que és una normal

07_T2_Interval_MitjanaPoblacional_cadamostra_DES.R

07_T2_Interval_MitjanaPoblacional_valoresperat_DES.R

ECET 2.6

Tema 2. Estimació puntual i per interval

INTERVAL DE CONFIANÇA PER A LA VARIÀNCIA POBLACIONAL (UNA MOSTRA)

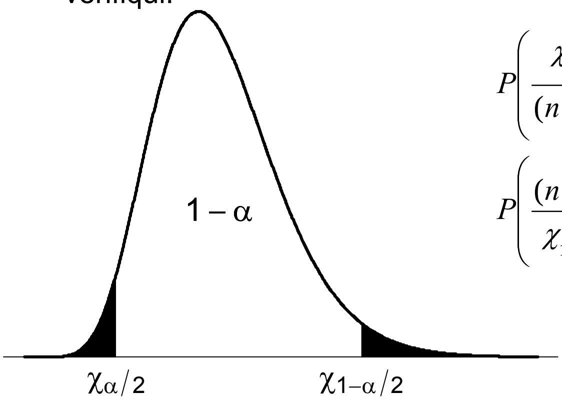
Sigui una població estadística caracteritzada per $X \approx N(\mu, \sigma)$

PRODEDIMENT

1.- Escollim el millor estimador de $\longrightarrow$ $\sigma^2 \rightarrow S_*^2 = \dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$

2.- Sabem, pel Tema 1 que: $\longrightarrow$ $\dfrac{(n-1)S_*^2}{\sigma^2} \approx \chi^2_{n-1}$

3.- Així, hem de construir un interval que ens permeti trobar els punts a la distribució que verifiqui: $\longrightarrow$ $\left(\chi^2_{\alpha/2,n-1} \le \dfrac{(n-1)S_*^2}{\sigma^2} \le \chi^2_{1-\alpha/2,n-1}\right) = 1-\alpha$



$$P\left(\frac{\chi^2_{\alpha/2\,n-1}}{(n-1)S_*^2} \le \frac{1}{\sigma^2} \le \frac{\chi^2_{1-\alpha/2\,n-1}}{(n-1)S_*^2}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)S_*^2}{\chi^2_{1-\alpha/2\,n-1}} \le \sigma^2 \le \frac{(n-1)S_*^2}{\chi^2_{\alpha/2\,n-1}}\right) = 1-\alpha$$

ECET 2.7

08_T2_Interval_VarianciaPoblacional_cadamostra.R

08_T2_Interval_VarianciaPoblacional_valoresperat.R

Tema 2. Estimació puntual i per interval

INTERVAL DE CONFIANÇA PER A LA PROPORCIÓ POBLACIONAL (UNA MOSTRA)

Sigui $\prod$ la proporció d'elements que compleixen una determinada característica. Atès que coneixem que el millor estimador és la proporció mostral, si la mida mostral és suficientment gran i es compleix que:

$n\Pi(1-\Pi) > 5$, aleshores $\rightarrow p \approx N\left(\Pi \le \sqrt{\dfrac{\Pi(1-\Pi)}{n}}\right)$

$$P\left(-Z_{\alpha/2} \le \frac{p-\Pi}{\sqrt{\dfrac{\Pi(1-\Pi)}{n}}} \le Z_{\alpha/2}\right) = 1-\alpha \quad \Rightarrow \quad P\left(p - Z_{\alpha/2}\sqrt{\frac{\Pi(1-\Pi)}{n}} \le \Pi \le p + Z_{\alpha/2}\sqrt{\frac{\Pi(1-\Pi)}{n}}\right) = 1-\alpha$$

Davant de la impossibilitat de fer servir Π per a poder construir l'interval, tenim dues opcions:

A.- Substituir Π per p:

$$P\left(p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \Pi \le p + Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

B.- Construir l'interval de màxima folgança, p = 0.5:

$$P\left(p - Z_{\alpha/2}\sqrt{\frac{0.5(1-0.5)}{n}} \le \Pi \le p + Z_{\alpha/2}\sqrt{\frac{0.5(1-0.5)}{n}}\right) = 1-\alpha$$

NOTES:

* Interval centrant en p

* Amplitud: $2 * Z_{\alpha/2}\sqrt{\dfrac{p(1-p)}{n}}$

* Error d'estimació: $d = Z_{\alpha/2}\sqrt{\dfrac{p(1-p)}{n}}$

09_T2_Interval_Proporcio_cadamostra.R 09_T2_Interval_Proporcio_valoresperat.R

ECET 2.8

Tema 2. Estimació puntual i per interval

INTERVAL DE CONFIANÇA PER A LA DIFERÈNCIA DE MITJANES POBLACIONALS (DUES MOSTRES)

Donades dues poblacions normals $X \approx N(\mu_x, \sigma_x)$ i $Y \approx N(\mu_y, \sigma_y)$

Sabem que $\overline{X} \approx N\left(\mu_x, \sigma_x \big/ \sqrt{n_x}\right)$ i $\overline{Y} \approx N\left(\mu_y, \sigma_y \big/ \sqrt{n_y}\right)$

i aleshores la variable aleatòria $\overline{X} - \overline{Y} \approx N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right)$

Ara podem establir l'interval de confiança per aquesta nova variable.

CAS A VARIANCES POBLACIONALS CONEGUDES

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \approx N(0,1) \quad \Rightarrow \quad P\left((\overline{X} - \overline{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \le \mu_x - \mu_y \le (\overline{X} - \overline{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right) = 1 - \alpha$$

NOTES:

* Centrat a $\overline{X} - \overline{Y}$ * Amplitud: $2 * Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$

* Si $\sigma_x^2 = \sigma_y^2$ idem tret que $\sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$ * Si X i Y no són Normals però $n > 30$

ECET 2.9

Tema 2. Estimació puntual i per interval

CAS B VARIANCES POBLACIONALS DESCONEGUDES PERO IGUALS

Atès que $\sigma_x^2 = \sigma_y^2 = \sigma^2$, però desconeguda, hem de proposar una estimació a partir de les dues mostres que tenim. Aquesta és $S_*^2 = \frac{(n_x - 1)S_{*x}^2 + (n_y - 1)S_{*y}^2}{n_x + n_y - 2}$. Però ara l'estandarització de la variable com que hem fet servir una estimació no segueix una distribució N(0,1)

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{S_*\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \approx t_{n_x + n_y - 2}$$

$$P\left((\overline{X} - \overline{Y}) - t_{\alpha/2, n_x + n_y - 2} S_* \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \le \mu_x - \mu_y \le (\overline{X} - \overline{Y}) + t_{\alpha/2, n_x + n_y - 2} S_* \sqrt{\frac{1}{n_x} + \frac{1}{n_x}}\right) = 1 - \alpha$$

NOTES:

* Centrat a $\overline{X} - \overline{Y}$ * Amplitud: $2 * t_{\alpha/2, n_x + n_y - 2} S_* \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$

* Per a mostres grans $S_*^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right) \approx \frac{S_{*x}^2}{n_x} + \frac{S_{*y}^2}{n_y}$

ECET 2.10

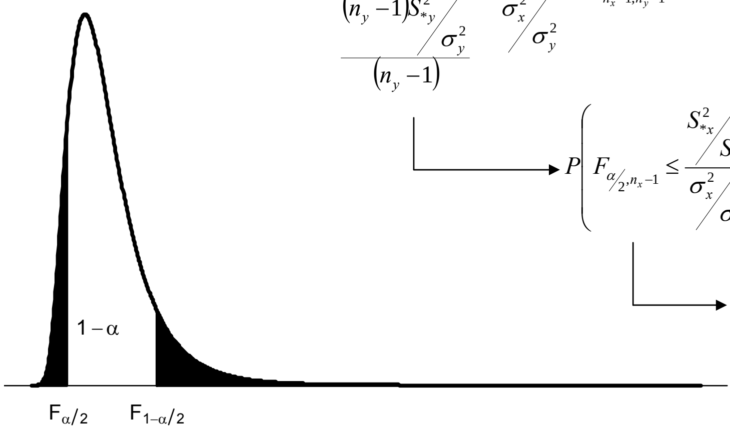
Tema 2. Estimació puntual i per interval

INTERVAL DE CONFIANÇA PER A LA DIFERÈNCIA DE VARIÀNCIES POBLACIONALS (DUES MOSTRES)

Donades dues poblacions normals $X \approx N(\mu_x, \sigma_x)$ i $Y \approx N(\mu_y, \sigma_y)$

Sabem que $\frac{(n_x-1)S_{*x}^2}{\sigma_x^2} \approx \chi^2_{n_x-1}$ i $\frac{(n_y-1)S_{*y}^2}{\sigma_y^2} \approx \chi^2_{n_y-1}$. També sabem que el quocient de dos distribucions $\frac{\chi^2_a / a}{\chi^2_b / b}$ es distribuiex segons una f - Snedecor : $F_{a,b}$. Per tant :

$$\frac{\dfrac{(n_x-1)S_{*x}^2 / \sigma_x^2}{(n_x-1)}}{\dfrac{(n_y-1)S_{*y}^2 / \sigma_y^2}{(n_y-1)}} = \frac{S_{*x}^2 / S_{*y}^2}{\sigma_x^2 / \sigma_y^2} \approx F_{n_x-1,n_y-1}$$



$$\longrightarrow P\left(F_{\alpha/2, n_x-1} \le \frac{S_{*x}^2 / S_{*y}^2}{\sigma_x^2 / \sigma_y^2} \le F_{1-\alpha/2, n_x-1} \right) = 1-\alpha$$

$$\longrightarrow P\left(\frac{S_{*x}^2 / S_{*y}^2}{F_{1-\alpha/2, n_x-1}} \le \sigma_x^2 / \sigma_y^2 \le \frac{S_{*x}^2 / S_{*y}^2}{F_{\alpha/2, n_x-1}} \right) = 1-\alpha$$

ECET 2.11

Tema 2. Estimació puntual i per interval

INTERVAL DE CONFIANÇA PER A LA DIFERÈNCIA DE PROPORCIONS POBLACIONALS (DUES MOSTRES)

Sigui Π_x i Π_y les proporcions d'elements que compleixen unes determinades característiques. Atès que coneixem que els millor estimadors són les proporcions mostrals, si les mides mostrals són suficientment grans i es compleix que:

$$n_i \Pi_i (1-\Pi_i) > 5, \text{ aleshores } \rightarrow p_i \approx N\left(\Pi_i \le \sqrt{\frac{\Pi_i(1-\Pi_i)}{n}} \right) \quad \text{per } i = \{x, y\}$$

$$P\left(-Z_{\alpha/2} \le \frac{(p_x - p_y) - (\Pi_x - \Pi_y)}{\sqrt{\dfrac{\Pi_x(1-\Pi_x)}{n_x} + \dfrac{\Pi_y(1-\Pi_y)}{n_y}}} \le Z_{\alpha/2} \right) = 1-\alpha \quad \Rightarrow \quad P\left((p_x - p_y) \pm Z_{\alpha/2} \sqrt{\frac{\Pi_x(1-\Pi_x)}{n_x} + \frac{\Pi_y(1-\Pi_y)}{n_y}} \right) = 1-\alpha$$

Davant de la impossibilitat de fer servir Π_x i Π_y per a poder construir l'interval :

$$P\left((p_x - p_y) \pm Z_{\alpha/2} \sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}} \right) = 1-\alpha$$

ECET 2.12

2.6.- Determinació de la mida mostral (RECORDAR PRIMERA PART DEL TEMA)

L'estimació per interval s'ha presentat com una solució a l'error d'estimació. Aquest es desconeix des del moment en que es disposa de la realització mostral, atès que aquest no es pot comparar amb el paràmetre poblacional.

UN ALTRE SOLUCIÓ al mateix problema és determinar, prèviament, l'ERROR MÀXIM que es permetrà amb UNA PROBABILITAT PREFIXADA i, en funció d'aquestes restriccions DETERMINAR QUIN SERIA LA MIDA DE LA MOSTRA NECESSÀRIA.

A.) POBLACIÓ INFINITA.

A.1) Determinació de la mida per a la MITJANA POBLACIONAL. VARIÀNCIA CONEGUDA.

Sigui $X \approx N(\mu,\sigma)$ amb σ^2 coneguda i n la mida mostra a determinar :

$$X \approx N(\mu,\sigma) \rightarrow \overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow P\left(\mu - d \le \overline{X} \le \mu + d\right) = 1-\alpha \text{ on } d \text{ l'ERROR que es PERMET}$$

$$P\left(\frac{\cancel{\mu} - d - \cancel{\mu}}{\frac{\sigma}{\sqrt{n}}} \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\cancel{\mu} + d - \cancel{\mu}}{\frac{\sigma}{\sqrt{n}}}\right) = 1-\alpha \rightarrow P\left(\frac{-d\sqrt{n}}{\sigma} \le Z \le \frac{+d\sqrt{n}}{\sigma}\right) = 1-\alpha$$

$$\downarrow \qquad \downarrow$$

$$-Z_{\alpha/2} \qquad Z_{\alpha/2} \rightarrow Z_{\alpha/2} = \frac{+d\sqrt{n}}{\sigma} \rightarrow n = \left(\frac{Z_{\alpha/2}\sigma}{d}\right)^2$$

A.2) Determinació de la mida per a la MITJANA POBLACIONAL. VARIÀNCIA DESCONEGUDA.

Sigui $X \approx N(\mu,\sigma)$ amb σ^2 DESCONEGUDA i n la mida mostra a determinar :

$$X \approx N(\mu,\sigma) \rightarrow \overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow P\left(\mu - d \le \overline{X} \le \mu + d\right) = 1-\alpha \text{ on } d \text{ l'ERROR que es PERMET}$$

$$P\left(\frac{\cancel{\mu} - d - \cancel{\mu}}{\frac{\sigma}{\sqrt{n}}} \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\mu + d - \cancel{\mu}}{\frac{\sigma}{\sqrt{n}}}\right) = 1-\alpha \rightarrow P\left(\frac{-d\sqrt{n}}{S_*} \le \frac{\overline{X}-\mu}{\frac{S_*}{\sqrt{n}}} \le \frac{+d\sqrt{n}}{S_*}\right) = 1-\alpha$$

$$\downarrow \qquad \downarrow$$

$$-t_{\alpha/2,n-1} \qquad t_{\alpha/2,n-1} \rightarrow t_{\alpha/2} = \frac{+d\sqrt{n}}{s} \rightarrow n = \left(\frac{t_{\alpha/2,n-1}s_*}{d}\right)^2$$

A.3) Determinació de la mida per a la PROPORCIÓ POBLACIONAL

Sigui $p \approx N\left(\Pi, \sqrt{\frac{\Pi(1-\Pi)}{n}}\right) \rightarrow P(\Pi - d \le p \le \Pi + d) = 1-\alpha$ on d l'ERROR que es PERMET

$$P\left(\frac{\Pi - d - \Pi}{\sqrt{\frac{\Pi(1-\Pi)}{n}}} \le \frac{p-\Pi}{\sqrt{\frac{\Pi(1-\Pi)}{n}}} \le \frac{\Pi + d - \Pi}{\sqrt{\frac{\Pi(1-\Pi)}{n}}}\right) = 1-\alpha \rightarrow P\left(\frac{-d}{\sqrt{\frac{\Pi(1-\Pi)}{n}}} \le Z \le \frac{+d}{\sqrt{\frac{\Pi(1-\Pi)}{n}}}\right) = 1-\alpha$$

$$\downarrow \qquad \downarrow$$

$$-Z_{\alpha/2} \qquad Z_{\alpha/2} \rightarrow n = \frac{Z^2_{\alpha/2}\Pi(1-\Pi)}{d^2}$$

$\hat{\Pi} = 0.5$ o una estimació de Π que consideris raonable

Tema 2. Estimació puntual i per interval

B.) POBLACIÓ FINITA.

B.1) Mitjana amb variància coneguda $\rightarrow n = \dfrac{N * Z^2_{\alpha/2} * \sigma^2}{d^2 * (N-1) + Z^2_{\alpha/2} * \sigma^2}$

B.2) Mitjana amb variància DESCONEGUDA $\rightarrow n = \dfrac{N * t^2_{\alpha/2} * S^2_*}{d^2 * (N-1) + t^2_{\alpha/2} * S^2_*}$

B.3) Proporció $\rightarrow n = \dfrac{N * Z^2_{\alpha/2} * p(1-p)}{d^2 * (N-1) + Z^2_{\alpha/2} * p(1-p)}$

ECET 2.13